ENV514 Machine Learning in Environmental Science

2025 Fall (08/25/2025 - 11/25/2025)
MoWe 4:40PM - 5:55PM

Instructor

Dr. Liyin He, Ph.D. (Pronouns: Lee-yin HUH)
liyin.he@duke.edu
Office Hours: TBD
Office Location: LSRC A207

Bio: Dr. Liyin He's research focuses on understanding the relationship between ecosystem, air, and water in the face of climate change. She is particularly passionate about translating fundamental ecological knowledge into applied practices that deliver socioeconomic benefits. These include, but are not limited to, (1) enhancing forest management to increase carbon sequestration, (2) improving crop yields to support food security, and (3) mitigating urban heat to promote public health. To address these challenges, she employs three primary approaches: (1) spatiotemporal data analysis to interpret field and satellite observations, (2) machine learning techniques to generate novel datasets, and (3) process-based ecosystem models to uncover underlying mechanisms. (More details in Research Website)

Teaching Assistant

Shiqi Miao (Pronouns: SHIH-chee MEE-ow)
shiqi.miao@duke.edu
Office Hours: TBD
Office Location: To be scheduled each week according to classroom availability in LSRC/Grainger Hall

Bio: Shiqi is interested in addressing interdisciplinary challenges at the nexus of climate change, terrestrial ecosystems, urban thermal dynamics, and public health. By integrating remote sensing, modeling, and data science approaches, she seeks to advance scientific understanding and inform practical strategies for urban climate resilience.

What is this course about?

This course explores how to work with environmental datasets, which contain both temporal and spatial information, and apply foundational, widely used machine learning algorithms to better understand and address environmental challenges. Through real-world case studies and

hands-on coding, we will use modern data-driven tools to uncover patterns in complex environmental systems.

What background knowledge do I need before taking this course?

To be successful in this course, students should have:

- Basic programming experience, preferably in Python
- Familiarity with data analysis, such as working with datasets, plotting, or running simple statistics
- Prior coursework in environmental science, ecology, or climate science is helpful but not required

What will I learn in this course?

Upon the completion of this course, you will be able to:

- Explain the use of machine learning in environmental science
- Apply supervised and unsupervised machine learning methods (e.g., regression, classification, clustering) to real environmental datasets
- Evaluate model performance using appropriate metrics and clearly interpret the results in the environmental context

What will we do in this course?

To help you achieve these objectives, we will engage in a mix of:

- Hands-on coding assignments using real environmental datasets (e.g., remote sensing observations, flux towers data, climate reanalysis data)
- Reading and discussing scientific papers to understand how machine learning is applied in real-world environmental research
- A final project, where you will apply what you have learned to a dataset and research question of your choice

How can I prepare for the class sessions to be successful?

To get the most out of this course, students should plan to engage both before and during class.

- Before class: You will occasionally be asked to complete short readings or watch brief videos. These materials will introduce key concepts and help you come to class prepared for hands-on work and discussion.
- During class: Sessions will be interactive and focused on applying what you have learned through coding exercises, group discussions, and collaborative data exploration. You are encouraged to bring your questions, share your insights, and help each other troubleshoot code.

- After class: You will reinforce your understanding through homework assignments, which are related to what we cover in class.

What are the required texts and materials?

Some useful coding tutorials and textbooks are listed:
https://jakevdp.github.io/PythonDataScienceHandbook/
https://developers.google.com/edu/python/introduction
Machine Learning for Earth Sciences: https://link.springer.com/content/pdf/10.1007/978-3-031-35114-3.pdf

If you are having difficulty with textbook, software, or supply costs associated with this course, here are some resources:

- Contact the financial aid office (whether you are on aid or not). They have loans and resources for connecting students with programs on campus that might be able to help alleviate these costs.
- DukeLIFE offers course materials assistance for eligible students. Please note that students who are eligible for DukeLIFE benefits are notified prior to the start of the semester; program resources are limited. Students who have limited access to computers may request loaner laptops through the DukeLIFE Technology Assistance Program. Please note that supplies are limited.
- Duke Libraries offers textbook rentals through the Top Textbooks at the Duke Libraries, where you can rent a textbook for 3 hours at a time.

How will my learning be assessed, and how will my grade be determined?

Your final grade in this course will be based on a combination of assignments, projects, discussions, and participation. These components are designed to reinforce learning, encourage engagement, and give you multiple ways to demonstrate your understanding.

Grading Breakdown:

Assessment	Weight	Description
Participation	15%	Active engagement in class, including attendance (10%), in-class discussion and peer feedback (5%).
Homework Assignments	45%	Regular exercises to help you understand machine learning concepts, practice data analysis techniques.
Paper Discussions	15%	Reading and discussing scientific papers that connect machine learning with environmental science applications.
Final Project	25%	A project in which a group of students will select a topic and dataset, apply machine learning methods and present findings.

Active Participation (15%) includes attendance, contributions to discussion, and peer feedback:

- Attendance (10pts)
  - ≤ 2 unexcused absences: 10 points
  - 3 unexcused absences: 8 points
  - 4 unexcused absences: 6 points
  - 5+ unexcused absences: 0–6 points (at instructor's discretion)
- In-class Discussion & Peer Feedback (5pts)
  - Excellent (5 pts): Consistently contributes to discussions and provides thoughtful, constructive feedback to peers.
  - Good (3–4 pts): Occasionally participates and offers some feedback; generally attentive and engaged.
  - Limited (0–2 pts): Rarely contributes to discussions and provides little or no peer feedback.

Homework Assignments (4 in total, 45%):

- Analysis of Tabular and Spatiotemporal Data (10 pts): Basic visualization and analysis of tabular and spatiotemporal datasets.
- Linear and Logistic Regression Practice (10 pts): Apply linear and logistic regression models to environmental data and interpret the results.
- Ensemble Methods and Explainable AI (15 pts): Implement Bagging and Boosting algorithms to analyze environmental dataset and use explainable AI tools to interpret model predictions.
- Neural Networks Practice (10 pts): Train and evaluate neural network models using environmental data.

Here are sample papers (link) you may consider for Paper Discussions. You are also encouraged to suggest any paper that interests you, including overview articles, perspectives, or case studies. Your presentation may cover the following key points:

- **Introduce the authors** (briefly): Who are the authors? What do we know about their lab?
- **Research background and objectives**: What has already been done in this area? What background are the authors building on? What is the goal of the paper, and why is it interesting and important?
- **Methodology**: How did the authors design the experiments, and how did they apply the machine learning methods. If some parts were hard to understand, take this opportunity to clarify them—use the board or additional materials to explain.
- **Results**: Present the results and help the audience focus on what is most interesting and important, especially if the paper contains many plots or data.
- **Discussion**: Share your thoughts about the paper—Are the authors addressing an interesting problem? How did they contribute to this field? Can this paper inspire new ideas on your project?

You don't have to follow this exact order. Feel free to go beyond these questions—use them as starting point to organize your thoughts. You are encouraged to ask the audience probing and bold questions to spark the discussion!

Paper discussion (15%) will be assessed in three parts:

- 6% – Comprehensive and clear introduction to the methods and results of the paper
- 5% – Critical thinking about and ideas inspired by this paper
- 4% – Presentation performance, i.e., speed, body language, clarity, and on-site discussion

Final project (25%) will be assessed in three parts:

- 5% – Project proposal (template link, you may also refer to the course documents)
- 10% – Oral presentation
- 10% – Project paper

Grading Scale:

- **A+** = 97–100
- **A** = 93–96
- **A–** = 90–92
- **B+** = 87–89
- **B** = 83–86
- **B–** = 80–82
- **C+** = 77–79
- **C** = 73–76
- **C–** = 70–72
- **D** = 60–69
- **F** = below 60

Late Submission Policy:

You are allowed one late submission of assignments of up to *1 day* past the deadline without prior approval. Just be sure to notify the TA in advance. After that:

- Late submissions will incur a 10% grade deduction in the corresponding assignment per day, unless you have a valid reason (e.g., medical or personal emergency).
- No late submissions will be accepted for the final project report.

What is the course schedule?

Week/Date	Topics & Activities (M)	Topics & Activities (W)	Assignments & Due dates
1 (8/25, 27)	Course intro: • Course overview • Student/instructor introduction	Lecture: • ML applications in ES overview	

2 (9/1, 3)	Labor Day (no class)	Lecture + In-Class Exercise: • Coding environment: Jupyter Notebooks • Set up Duke Compute Cluster (DCC)	
3 (9/8, 10)	Lecture: • Data types (tabular, raster, vector) & sources • Python basics	Assignment #1: • Tabular data visualization and analysis	Assignment #1: Out 9/10, due 9/19
4 (9/15, 17)	Assignment #1: • Spatiotemporal data visualization and analysis	Lecture + Exercise: • Linear regression	
5 (9/22, 24)	Lecture + Exercise: • Support Vector Machine • Logistic regression	Assignment #2: • Linear regression • Logistic regression	Assignment #2: Out 9/24, due 10/3
6 (9/29, 10/1)	Lecture and Exercise: • Decision Trees • Bagging (e.g., Random Forest) • Boosting	Lecture + Exercise: • Feature Importance • Explainable AI with SHAP	Paper Selection: Submit title and link via provided Google Sheet by 10/3
7 (10/6, 8)	Guest Lecture: Explainable AI and causal analysis	Work on final project proposal.	Final Project Proposal: Due 10/15
8 (10/13, 15)	Fall Break (no class)	Assignment #3: Applications of bagging, boosting and explainable AI	Assignment #3: Out 10/15, due 10/24
9 (10/20, 22)	*Instructor and TA will meet with each group to provide feedback on the proposal*	Guest Lecture: • AI + remote sensing in Google Earth Engine	
10 (10/27, 29)	Paper Discussion I	Paper Discussion II	
11 (11/3, 5)	Lecture + Exercise: • Unsupervised learning • K-means	Lecture + Exercise: • Neural Networks	
12 (11/10, 12)	Assignment #4: • Neural Networks	Work on final project	Assignment #4: Out 11/10, due 11/21
13 (11/17, 19)	Work on final project	Final project presentation I	

14 (11/24)	Final project presentation II (*tentative, depending on how many groups we have)*	/	Final Project Report: Due 12/10

Course Communication

Course announcements, including schedule changes, class cancellations (e.g., due to weather or illness) and assignment updates, will be communicated via Canvas Announcements or email. If you are ill or experiencing a personal emergency and are unable to submit an assignment on time, please notify both the TA and me as early as possible, ideally before the assignment is due. We are happy to work with you to make accommodations when needed. If you have questions or concerns about course content, assignments, or logistics, you are welcome to email me directly or visit during my office hours. For questions specifically related to grading or assignment feedback, please contact the TA directly or attend TA office hours. I will typically respond to emails within 48 hours on weekdays. Messages sent over the weekend will typically be addressed the following Monday.

Discussion Guidelines

Civility is a fundamental component of productive academic discourse. In this course, we welcome a diversity of ideas, experiences, and perspectives. Active, respectful interaction with your peers and instructor is key to your success. Please keep the following expectations in mind when engaging in class discussions and activities:

- Respect differing viewpoints. It is okay to disagree, but critique ideas—not individuals.
- Listen actively and speak considerately. Avoid interrupting and be mindful of tone and body language.
- Support your contributions with evidence, examples, or clear reasoning.
- Be mindful of airtime. If you tend to speak often, make space for others. If you are more reserved, we encourage you to find opportunities to share your perspective.

If you witness or experience communication that you believe violates these guidelines, please bring it to the instructor's attention. We are committed to fostering a respectful and inclusive learning environment for all.

What are the course policies?

Academic Standards:

All students must adhere to the Duke Community Standard (DCS): Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and

uphold these principles in all academic and nonacademic endeavors, and to protect and promote a culture of integrity.

To uphold the Duke Community Standard, students agree:

- I will not lie, cheat, or steal in my academic endeavors.
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised.

Regardless of course delivery format, it is the responsibility of the student to understand and follow Duke policies regarding academic integrity, (e.g., completing one's own work, following proper citations of sources, and adhering to guidance around group work, and more). Ignoring these requirements is a violation of the Duke Community Standard. Any questions and/or concerns regarding academic integrity can be directed to the Office or Student Conduct and Community Standards at conduct@duke.edu.

AI Policy:

In this course, you are welcome to use AI tools—such as ChatGPT, GitHub Copilot, Gemini, and others—to support your learning process. These tools can be helpful for idea generation, debugging, and interpreting code, and may help you explore concepts or troubleshoot issues as you work through assignments. However, to ensure you fully engage with the course material and to maintain academic integrity, it is essential that all submitted work be entirely your own. Any content you submit—whether written explanations, analysis, or code—must be your original work and you are responsible for explain your code and written reports.

Academic Accommodations:

If you are a student with a disability and require accommodations for this course, it is your responsibility to register with the Student Disability Access Office (SDAO) and provide appropriate documentation. SDAO will work with you to determine reasonable accommodations based on your individual needs. Please note that accommodations cannot be implemented until the instructor receives an official Faculty Accommodation Letter from SDAO. I encourage you to begin this process as early as possible to ensure timely support. For more information or to begin the registration process, contact SDAO at sdao@duke.edu

Mental Health and Wellness Resources

Your mental health and well-being are a top priority at Duke. The university offers a variety of resources to support students in managing stress, maintaining balance, and taking care of their overall well-being. If you are feeling overwhelmed, please know that you are not alone, and help is available. Some resources are listed below:

DuWell provides Moments of Mindfulness (stress management and resilience building) and meditation programming (Koru workshop) to assist students in developing a daily emotional well-being practice. All are welcome and no experience is necessary.

If your mental health concerns and/or stressful events negatively affect your daily emotional state, academic performance, or ability to participate in your daily activities, many resources are available to help you through difficult times.

DukeReach provides services for students who are experiencing significant challenges relating to mental health, physical health, social adjustment, and/or a variety of other stressors.

Counseling & Psychological Services (CAPS) services include individual and group counseling services, psychiatric services, and workshops. To initiate services, walk-in/call-in 9:00 AM – 4:00 PM (M/W/Th/F) and 9:00 AM – 6:00 PM Tuesdays. CAPS also provides referral to off-campus resources for specialized care. Contact: (919) 660-1000

TimelyCare (formally known as Blue Devils Care) is an online platform that is a convenient, confidential, and free way for Duke students to receive 24/7 mental health support through TalkNow and scheduled counseling.